DOCUMENT RESUME

ED 269 479 TM 860 339

AUTHOR Rose, Janet S.; Gustin, William C.
TITLE Elementary/Middle School Reading Comprehension vs. High School Mathematics: Two Different Approaches to CRT Development.
PUB DATE Apr 86
NOTE 28p.; Paper presented at the Annual Meeting of the American Educational Research Association (70th, San Francisco, CA, April 16-20, 1986). For related document, see TM 860 336.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Criterion Referenced Tests; Curriculum Development; Educational Objectives; Educational Testing; Elementary Secondary Education; *Mathematics Tests; Measurement Objectives; Pilot Projects; *Reading Comprehension; Reading Tests; School Districts; Task Analysis; *Test Construction; Test Items
IDENTIFIERS *Charleston County School District SC; *Test Specifications

ABSTRACT

This paper details the development of the high school general mathematics examination and reading comprehension tests for grades one through eight for the curriculum referenced testing program in Charleston County School District, South Carolina. While the basic developmental approach was similar, problems encountered in developing the two types of tests were unique and inspired different strategies to accommodate the differences in the curricula. Begun in 1983 development of the ninth grade mathematics examination involved identifying objectives and writing item specifications, followed by four cycles of item writing, test production, and pilot-testing. Curriculum changes by the North Office in 1983 necessitated recalibrating 468 items. After pilot testing a second set of test forms in 1984, a consulting firm recalibrated the item bank, created four test forms and projected student performance. The first official examination administration was postponed until 1986 and the 1985 administration became a field test. The examination will account for 50 percent of the final course grade. The test development process goal for reading comprehension was to generate formative and summative tests to assess curricular objectives for grades one through eight. Outside consultants assisted in identifying reading comprehension objectives due to the different organization of existing elementary and middle school objectives. Outside contractors were also used to develop test specifications and to train district teachers in writing the test items. Item review and revision by district staff took longer than anticipated, so a language arts content expert completed the item review and prepared the 60 pilot test forms. Pilot testing was postponed until the spring of 1986. (BS)

ED269479

ELEMENTARY/MIDDLE SCHOOL READING COMPREHENSION VS. HIGH SCHOOL MATHEMATICS: TWO DIFFERENT APPROACHES TO CRT DEVELOPMENT

Janet S. Rose
Charleston County (S.C.) School District

William C. Gustin
Charleston County (S.C.) School District

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. S. Rose

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM 860 339

AERA Symposium: Problems and Practical Solutions in the Development of Districtwide Curriculum-Referenced Tests

San Francisco
April, 1986

Charleston County's curriculum-referenced testing program began with the development of an examination for high school general mathematics. Another major effort has been the development of reading comprehension tests for grades one through eight. While the basic developmental approach was similar, problems encountered in the development of the two types of tests were unique and inspired different strategies. The following case studies highlight our efforts to accommodate and even capitalize upon differences in the curricula.

GENERAL MATHEMATICS I

Test Development Bulletin number 2, produced by the Office of Evaluation and Research and dated June 30, 1982, contains the following announcement: "As part of the test development scheduled for the next few years, countywide area examinations for high school courses are planned. Area examinations will assess content specific to a particular course. The examinations will consist of teacher-written items for course-specific objectives." The Test Advisory Committee decided that priority be given to math and language arts, beginning with grade 9 and moving up the grade levels. Therefore, ninth grade General Mathematics was chosen as the first area examination to be developed. Additionally, a study guide for this course had been distributed to teachers and the curriculum was thought to be stable.

Our early efforts were permeated with a certain sense of urgency. We were asked to produce semester examinations based upon item pilot tests administered at the end of each nine-week period. Chapter II funding was secured and work began immediately with an assessment of the General Mathematics objectives. At the time, the content of the course was comprised of 133 instructional objectives, of which 94 were designated as "core."

IDENTIFICATION OF OBJECTIVES

Initially we felt an obligation to address the notion of objective "mastery." We were advised that an estimate of mastery would require at least

six items per objective. We noted that six items multiplied by 133 objectives translated into a least 700 more items than the longest test we had considered giving. Clearly it would be necessary to take a hard look at the list of objectives. The first of several small groups of teachers was convened and asked to classify each objective as an "essential" or "less essential" element of course content, or as wholly "subsumed" by another objective. The exercise pared our list of objectives down to 99. At that point we began discussing the possibility of assessing clusters of objectives, that is, content "domains."

TEST AND ITEM SPECIFICATIONS

The following year was a whirlwind of activity. A second group of teachers agreed upon an item specification form and wrote specifications for each of the 99 objectives. The objectives were reviewed by district personnel and experts in math education from a local college and revised accordingly. The objectives were organized into four groups according to which quarter they were most likely taught.

ITEM WRITING

Near the beginning of the school year, an additional staff member was hired to coordinate test development. At the same time yet another group of teachers was recruited and asked to write a total of 20 items for each of the first-quarter objectives. We had been advised that roughly half of the items would not survive a critical review and possibly one-fourth of the remaining items would be rejected after the item pilot. An item-writing workshop was organized for the teachers who wrote items. The specification form was explained to them and they were indoctrinated with the do's and don'ts of the multiple-choice format.

Five teachers and one representative of the district's math office wrote a total of 300 items over a two-day period. At the time it was still possible

to pull teachers from their classes, so the only expense involved was to reimburse the general substitute account.

An item-review form was devised to facilitate a formal review and focused upon a variety of potential flaws and a consideration of validity and bias. A design for the first quarter item pilot was suggested by a technical consultant: nine forms of a 36-item test with pairwise linking between the forms. The design required 162 items, nine items for each of the eighteen objectives. Based upon the completed review forms, a preliminary sort of the items was conducted by E&R staff. If any reviewer noted a flaw or questioned the validity of an item, the item was rejected. If more than our goal of nine items per objective survived the review process, items were selected to represent a variety of styles and difficulties. If less than nine items survived, E&R staff evaluated the rejected items and, based upon the reviewer's comments, revised enough items to fill the quota.

FIRST ITEM PILOT

Test booklets were produced, instructions written for students, and a checklist prepared to aid teachers in the administration of the pilot test, which involved 50 teachers and over 2,000 students. After the test was given, all materials were returned to our office and the answer sheets were prepared for scanning. A scanning and editing program was written for use with an "antique" IBM 5100 microcomputer and a 3M scanner. Two graduate assistants were taught the scanning and editing procedures. Student records were entered onto tape cassettes and, when the scanning was complete, all data were transmitted to mass storage at the University of South Carolina. From there a consultant would access the file for the item analysis.

This cycle of item writing, pilot test production, administration, and scanning was repeated three more times by the end of the school year. A consultant was retained throughout that year to conduct a Rasch item analysis

and calibrate the bank of general math items. During that first year, support shifted away from the two semester exams in favor of one final examination. Each of the four pilot tests was linked by a sufficient number of items to allow the calibration of the final bank as if all of the items had been administered at the end of the year. That bank contained a total of 793 items.[1]

Instructional survey. As our test development efforts began, the attention of the testing community was focused upon the legal challenge to Florida's testing program. Our response was to proceed with caution and incorporate checks and balances and significant teacher involvement. We designed an Instructional Record Survey. Teachers were asked to indicate when they taught which objectives and how well those objectives were mastered by their students. The fourth pilot test was modified to include items from objectives which many teachers had not covered in time for the third pilot. Another survey was designed to give every General Math teacher an opportunity to react to the classification of the General Math objectives. Subsequently, 11 objectives were reclassified.

Item review. Content and bias reviews were conducted during the summer of 1983. A limited number of items was deleted from the bank because of concerns of possible bias. A formal content review was devised and confirmed the match between objectives and items. The judges were in agreement on over 98% of the items.

CURRICULUM CHANGE

Our first major setback occurred during that same summer, with the discovery that the Math Office had also been busy. The General Mathematics study guide had been revised. Sixteen objectives had been dropped, seventeen new objectives written, seventeen revised, and every objective renumbered. The impact of the changes essentially nullified one-third of the work we had

done our first year. Heated memos were exchanged and a series of meetings held. Once again the curriculum was scrutinized from the various perspectives of content, instruction, and evaluation. When the dust had settled, we were asked to recalibrate 468 items associated with 69 "essential," "core" objectives.

SECOND ITEM PILOT

In spite of the setback, we had learned a great deal during our first year of test development. Based upon the results of the first pilot, the bias review and an analysis of fit, a second set of item pilot test forms was designed and the tests administered in May of 1984. Concurrent with our efforts to resolve our problems with the General Math curriculum, we had proceeded with the development of item pilot tests for mathematics at grades 2, 3, 5, 6, and 8. Those tests were also administered in May of 1984. It was obvious that we had outgrown our scanning capability. Consequently, new answer sheets were designed and the district's data processing office was enlisted to scan answer sheets on their highspeed scanner and create computer tapes on their mainframe computer.

A test development consulting firm was contracted to analyze the pilot test results: to recalibrate the item bank, to create four parallel test forms, and to project student performance on the operational forms. One blessing of the second item calibration was the ability to verify the stability of the item difficulty estimates over time. Average item difficulties within content domains were found to remain stable--usually within one or two hundredths of a logit. The four test forms were designed to contain 77 items. Each test contains items from every content domain, weighted according to relative emphasis in the course, with at least one item from every core objective.[2]

PERFORMANCE STANDARDS

Our intent was to administer the first of the four parallel forms in May of 1985. A series of three meetings was held during the 1984-85 school year. Divisional directors, high school principals, and General Mathematics teachers were given the opportunity to review the first form item-by-item and to specify a raw score which would represent a minimum standard of performance. The average score selected by the three groups was 56 of 77. Projections based upon student performance on the second item pilot indicated that a standard of 56 would translate into a failure rate of 94%. The results of the standard-setting sessions were summarized in a memo to the Deputy Superintendent of Instruction along with a request for a decision regarding the disposition of the test.

The District Superintendent's administrative team decided to phase in the area exam gradually by postponing the first official administration of the exam until the spring of 1986. The 1985 administration was to be considered a field-test. The Director of Vocational Education was asked to explore the relevance of the course objectives to life and work skills. Our office was asked to report on the results of the 1985 field-test and to conduct another survey of teachers.

The average raw score in 1985 was six points higher than expected, probably due to the increase in seriousness with which the test was taken. Even though the test was not operational countywide, most teachers used it as their own final examination.

INSTRUCTION AND ASSESSMENT

A survey was conducted asking teachers to rate the difficulty of each course objective and to indicate whether or not they taught the objective. The correlation between objective difficulties based upon student performance and those indicated by teachers was over 0.7. And, the survey confirmed that

most of the teachers believed they were teaching each objective. A 29-page report containing these and other informative analyses was presented to the Superintendent in September of 1985. The policy he established for the implementation of the General Mathematics Area Examination includes the stipulation that the exam will account for 50% of the final course grade and establishes a grading scale for the exam such that the projected failure rate for the exam is 53%. We believe that the impact of the new policy is not as brutal as it may sound. Approximately 50% of all General Math students in Charleston County customarily fail the final exams prepared by their teachers. However, as 50% of the final grade, the exam should have a considerable impact on the consistency of grading across the county.

READING COMPREHENSION

The goal of the test development process for language arts is to generate a set of formative and summative tests to assess curricular objectives for grades 1-8. The district's language arts curriculum consists of four strands: reading comprehension, word recognition (grades 1-5 only), study skills, and composition. At the time test development began, the curriculum for grades 1-5 had teachers' guides for each strand. These guides contained the objectives, and for each objective, a set of instructional strategies and a three item "criterion-referenced test" to be used for assessment purposes. Since subject area staff chose reading comprehension as the first priority for test development, efforts began with a study of the existing language arts objectives at the elementary and middle school grade levels. Also included was an analysis of the guide's test items to determine their usability as formative assessment tools.

IDENTIFICATION OF OBJECTIVES

Several issues of concern surfaced as a result of studying the objectives and related curricular materials. A major finding was that elementary and

middle school objectives were developed by two different staff members. Consequently, the objectives for the two school levels were organized differently, as illustrated in Table 1. The elementary strand consisted of two broad domains--Literal Comprehension and Interpretive Comprehension. These domains contained five and nine "terminal" objectives, respectively. Subsumed beneath each terminal objective was a set of "process" objectives, arranged hierarchically by difficulty. These were considered to be prerequisites to attainment of the terminal objectives. For example, identifying details (a) heard in a story, (b) seen in a picture, or (c) read in a sentence could be considered prerequisites to the terminal objective, stating details in a reading selection. For some terminal objectives, for example, Word Meaning, the process objectives seemed to be subskills, e.g., identifying antonyms, synonyms, homonyms and multiple meanings. In contrast, the middle school reading comprehension curriculum contained three domains: Literal, Inference, and Analysis of Literature. Each domain contained a set of terminal and subsumed process objectives. These process objectives were primarily subskills of the terminal objective, though, occasionally, they were arranged hierarchically, as from "Identify statements that express the main idea" to "Identify paraphrased main ideas" to "Paraphrase the main idea of a reading selection." The chart of objectives indicated whether each objective was "instructed," "emphasized," or "maintained" at each grade level.

Review and revision of objectives. Evaluation and Research (E&R) staff asked if the Language Arts (LA) staff might want to consider creating a single organizational framework for the grade 1-8 objectives. The LA staff agreed that this was a curricular and instructional necessity, as well as an essential prerequisite to the development of a sequential set of tests for the grade 1-8 testing program. In addition, since they were generally

dissatisfied with some of the objectives, the LA staff chose to take this opportunity to revise the existing curricula.

A decision was made to employ two external consultants (one, an expert in reading, the other, an expert in measurement) to facilitate the preparation of reading comprehension objectives and to help ensure their grade level continuity and measurability. Each of the following problems was addressed at a two-day meeting attended by both the LA and E&R staff.

1. Organization of objectives--A major concern of the LA staff was that the end-of-year test provide teachers with objective-level mastery information. It was pointed out, however, that tests designed to provide teachers with student "mastery" information on individual objectives would be too long (assuming a minimum of six items per objective). It was recommended that objectives be grouped into domains, whereby a sufficient number of items could be included to provide teachers reliable mastery information at the domain level. Grouping objectives into only three domains, such as Literal, Inferential, and Analysis of Literature probably would not be useful to teachers. Therefore, it was decided that the terminal and related process objectives be grouped into domains. For each terminal process objective, the LA staff assigned grade levels in which it should be taught.

2. Leveling--This problem refers to the difficulties experienced by the LA staff in determining the grade levels at which objectives should be taught. Sometimes the problem was simply a question of when content should be taught, e.g., At what grade levels should antonyms or homonyms be taught and tested? At other times, the problem focused on the context in which an objective should be taught and tested, e.g., "Given a paragraph, the student will identify the main idea" vs. "Given a reading selection, the student will identify the main idea." It was suggested

that this latter concern be addressed by the test specifiers. The staff should concentrate their efforts on decisions regarding when content should be taught and the specifications of context only when the objectives involved different processes (e.g., identifying the main idea in a picture vs. in a story presented orally).

3. Assessment--There was a clear preference by the LA staff that the terminal objectives be written and tested in open-ended rather than multiple-choice fashion. For example, they preferred that the learner be able to "state" rather than "identify" the main idea, since the former was a more accurate reflection of reality. E&R staff described the logistical constraints of administering tests with open-ended items and explained to the LA staff that multiple-choice items could be written which would be as difficult as open-ended items. The LA staff insisted on the open-ended nature of the curriculum. Though a compromise could not be arranged for testing purposes, it was agreed that in the classroom the teacher could instruct and assess the objective in an open-ended way.

After a painful struggle, a set of curricular objectives finally emerged. Though the new reading comprehension curriculum contained some minor inconsistencies, it satisfied curriculum and measurement requirements. The present reading comprehension strand contains 14 terminal objectives with an additional 59 process/subskill objectives. The curriculum is printed in matrix form, with the objectives listed on the left side of the page and the grade levels across the top. X's are placed in the cells to indicate grade levels where terminal and process objectives should be taught.

Selection of objectives for summative tests. The final preliminary step to the development of tests was the identification of objectives to be tested on the end-of-year summative tests. This issue generated discussions concerning the purpose of the tests. The resolution for purpose was that the

tests should serve both diagnostic and evaluative functions. Therefore, objectives were selected if they were goals for that year, or if they could provide useful diagnostic information about students' current levels of functioning. Formative assessment devices would eventually be developed for objectives omitted from the summative tests so that teachers could assess students' progress throughout the curriculum. The number of objectives tested for each grade level ranged from 16 to 19. A section of the revised objectives chart appears in the Appendix. Objectives tested at the end of the year are circled.

Blueprinting. It was decided not to create a test blueprint at this time because (a) it was not necessary for the next few steps (i.e., creation of specifications and items), and (b) we felt that we could not accurately predict the number of items that should appear on a test form until we knew more about the specifications (e.g., length of reading passage).

TEST AND ITEM SPECIFICATIONS

Earlier, a review of existing reading comprehension items developed by the LA staff was initiated. An external consultant had been employed in order to obtain an unbiased evaluation of the items. This measurement expert advised us to begin anew. He cited several problems with the items. AMong them were inappropriate item format, inadequate use of visual stimuli (graphics), culturally-specific items, lack of content validity, inconsistent readability and great variability among items assessing an objective. He concluded his report by recommending that test specifications be written to ensure the quality of test items.[3]

The question of how test specification should be prepared involved three factors--financial cost, staff time and staff expertise. For several reasons the district chose to employ an outside contractor to coordinate specifications activities for this particular area. First, due to other

district responsibilities, central office staff were not able to devote the considerable amount of time needed to complete this task. Second, though the district had language arts and measurement experts on its staff, these individuals did not have any experience writing language arts specifications. Although the same was true for math, it was felt that expertise was needed to direct the construction of specifications in an area as difficult as reading comprehension.

A Request for Proposals (RFP) for development of test specifications was distributed. The RFP outlined the general procedures for how the task was to be accomplished, including the roles and responsibilities for district staff, teachers and contractor. Briefly, a committee of district staff and teachers would form a specifications committee to provide input into the detailed content of the specifications and to review all versions of the specifications. The contractor would also employ their own content experts to review the specifications.

E&R and LA staffs reviewed the proposals received in response to the RFP. The two staffs chose different contractors. The E&R staff chose an agency having considerable background and experience in test development, whereas the LA staff selected a company with a strength in language arts content. Each staff feared that selection of the other's choice would compromise their respective areas of concern. After much discussion, debate, and negotiation, a final decision was made to award the contract to the bidder chosen by E&R. The language arts expert associated with the other bidder agreed to participate in the development and review of the specifications as an external consultant.[4]

Prior to the initiation of specification activities, a planning meeting was held to discuss the development of a framework document. This document would define the test specifications format, address various issues and

describe how, when, and under whose responsibility each step of the project would be accomplished. Questions addressed in the framework were:

1. What components should be contained in the test specifications? Components included a general description of the objective, a sample item, a description of the test question, a description of the answer choices, and an optional content supplement.
2. At what level should the language arts objectives be specified? Domain level specifications were written to ensure curricular and instructional continuity across the grade levels. Differences among grade-level objectives within domains were addressed.
3. What difficulty level variables should be considered for each domain? These included readability, word level, sentence length, passage length, and number of answer choices.
4. What readability procedure should be selected? A comparison among the various options resulted in selection of the Fry formula.
5. What word list should be selected? The South Carolina Word List was used.
6. What type of content material should be used? All types were used--fiction, nonfiction, poetry, etc.
7. Should locally relevant content be used? No systematic attempt was made to use locally relevant content, though content which was unfamiliar or irrelevant to district students was excluded.
8. What is the most effective way for the test specifications committee to be used? Rather than arrange for teachers to work on specifications for their own grade level, committees composed of teachers from all grade levels worked on entire domains. This procedure would ensure grade level continuity with the domain-level approach to the specifications.

Once the above procedures were specified, two meetings were scheduled for the purpose of providing the contractor with input on test specifications.

The contractor facilitated the meetings which were attended by 20 teachers and the language arts content expert. The contractor used the information gathered at the meetings to create a preliminary version of the specifications which was sent to the district for review. At that time, another two-day meeting was scheduled to make the final adjustments to the specifications. The contractor, six of the original 20 teachers, and the language arts expert reviewed the specifications document. Specifications were revised accordingly by the contractor and sent to the E&R director and language arts coordinator for final review and approval. A reproduction of a specification (without sample items) is found in the Appendix.

ITEM WRITING

Again, the nature of the subject area led the district to believe that the skill and time required for item writing were more than what the central office could provide. Although, by this time, the district had had experience writing test items for mathematics, it was felt that not only would reading comprehension items be more difficult to write and time-consuming to review, they would cost much more in time and money in the long run if poor items had to be rewritten. Therefore, it was decided to distribute an RFP in the fall for the development of 146 reading comprehension items (10 items per objective) and for the design of the pilot-test forms to be administered in the spring.

A key feature of the proposal stipulated that district teachers would be trained and used as item writers. Use of district, as opposed to external, item writers would have three benefits. First, tests would be perceived by teachers as belonging to the school district and its teachers. This factor was particularly important in light of the two state-imposed testing programs currently operating in the district. Second, the procedure would ensure instructional relevance. Items would be based on content that is in accord

with instructional practices used by teachers. And, third, it was anticipated that the training and experience teachers received in item writing would carry over into the preparation of their own classroom tests.

Following award of the contract in December, a planning meeting was held with district staff and the contractor to clarify activities related to the development of test items and preparation of pilot forms.[4] A document that proved extremely useful in the development of test items was the domain item distribution plan prepared by the contractor. This document was written to ensure that items developed would sample the eligible content as representatively and completely as possible. The plan for each domain predefined key features of each item to be written, e.g., the nature of the reading selection used (such as fiction, nonfiction, poetry), the particular subskill, and the position of the correct answer, as well as the amount paid per item.

Item writing procedures began with a two-day training workshop conducted by the contractor. On the first day, potential teacher item-writers learned how to use test specifications, apply readability procedures, select content for items and develop high quality passages and answer choices. The contractor also discussed procedures associated with item writing assignments. The second day was devoted to actual item writing, with time provided for answering questions and reviewing items written that day. At the conclusion of the day, item assignments were made with the expectation that all items would be written within three weeks.

Teachers corresponded directly with the contractor during the item writing phase of the project. Teachers sent test items to the contractor; the contractor and contractor's staff reviewed these and forwarded final versions to district staff and our external language arts content expert for final review. This latter individual in turn forwarded his comments to the

district. All items and reviews were received in the district by mid-March, only two months prior to the anticipated pilot test.

Item review. Review of items by district staff began immediately with all-day meetings attended by selected district language arts staff and teachers. Time constraints prevented continuous scheduling of meetings, and, to our disappointment, the review process transpired more slowly than expected. Two all-day meetings resulted in a review of items for only 25 of the 146 objectives, (250 items or 17% of the total number of items). It was decided at this point to postpone the pilot-testing of the reading comprehension items until spring, 1986. The revised schedule included extention of item review meetings into the summer with the expectation that pilot-test forms would be prepared in the fall.

What caused the delay? First, it was unrealistic to expect that district staff could review 1,460 items and prepare and print pilot forms in a two-month period. Second, the contractor preferred to revise the items as little as possible in order to preserve "local" style. However, district staff felt that the items, being written by first-time item writers, needed closer scrutiny and revision. Therefore, district reviewers spent a lot of time making and remaking stylistic changes in the items. Third, the graphics for items requiring pictures were unacceptable and had to be redrawn.

The plan to complete the item review underwent another change. After several all-day item review meetings, it was decided that the task needed full-time attention. A language arts content expert was transferred from the language arts department to E&R for the purpose of completing the item review (revising and rewriting, if necessary), overseeing completion of new graphics by a local artist and preparing the 60 pilot-test forms. In addition, use of one full-time content expert would minimize stylistic differences among

reviewers. This individual has been working four days a week since October, under a great deal of pressure, to meet this spring's pilot-test dates.

PILOT-TESTING

The pilot-test design created by the contractor linked items horizontally within grades and vertically across grades by means of anchor items appearing on all forms within a grade and a special anchor form at each grade level containing lower and upper grade items. The addition of vertical linking was possible due to the across-grade sequence of objectives and the domain-level nature of the test specifications.

CONCLUSIONS

The development of the language arts reading comprehension items has been a learning experience for our district. We learned several lessons. One major lesson was that development of quality reading comprehension items takes more time and attention than we had expected. It also requires item writing skills that our district teachers did not have. In the end, however, the trade-off was worthwhile. Though we had to make many revisions to the test items, district teachers did gain experience in item writing and the tests will be viewed as belonging to the district. The costs were hiring a person to coordinate and complete item revisions and postponing the pilot-test for a year.

NOTES

[1]The consultant employed to analyze the first set of pilot-test forms was Dr. Joseph Ryan, Educational Measurement Systems, Inc.

[2]The contractor awarded the bid to analyze the second set of pilot-test forms was The Corporation for Measurement and Statistics (Paul Williams and Gary Phillips).

[3]The consultant employed to review the original language arts items was Dr. Joseph Ryan, Educational Measurement Systems, Inc.

[4]The contractor awarded the bids for the development of test specifications and test items was The Corporation for Measurement and Statistics. The individual within CMS responsible for preparing the specifications and items was Dr. Elaine Lindheim.

Table 1

Initial Organization of Reading Comprehension Objectives

Grades 1-5

	Literal	
CL	1.0	Word Meaning
CL	2.0	Detail
CL	3.0	Sequence
CL	4.0	Following Directions
CL	5.0	Main Idea (Stated)

	Interpretive	
CI	6.0	Main Idea (Inferred
CI	7.0	Relationships (Class)
CI	8.0	Relationships (Comparison/Contrast)
CI	9.0	Relationships (Cause/Effect)
CI	10.0	Relationships (Analogies)
CI	11.0	Predicting Outcomes
CI	12.0	Drawing Conclusions
CI	13.0	Figurative Language
*CE	14.0	Making Judgments

*E is Evaluation

Grades 6-8

	Literal	
MWA	1.0	Main Idea
MWM	2.0	Word Meaning
MLC	3.0	Details
MLC	4.0	Main Idea

	Inference	
MIC	5.0	Main Idea Inferred
MIC	6.0	Cause/Effect
MIC	7.0	Comparison/Contrast
MIC	8.0	Predicting Outcomes
MIC	9.0	Draw Conclusions
MIC	10.0	Analogies

	Analysis of Literature	
MAL	11.0	Figurative Language
MAL	12.0	Making Judgments
MAL	13.0	Story Elements
MAL	14.0	Rhetorical Devices
MAL	15.0	Fiction
MAL	16.0	Author's Purpose
MAL	17.0	Non-Fiction
MAL	18.0	Poetry
MAL	19.0	Plays

APPENDIX

- COMPREHENSION OBJECTIVES -
Grades 1 - 12
1984

OBJECTIVES	GRADE LEVELS											
	1	2	3	4	5	6	7	8	9	10	11	12
CI 9.0 RELATIONSHIPS (CAUSE/EFFECT) - The learner can identify implied causal relationships in a reading selection.				(X)	(X)	(X)	(X)	(X)				
9.1 The learner can identify cause and effect relationships by matching pictures.	(X)	X										
9.2 The learner can identify the causal relationship in a story presented orally.		X	X									
9.3 The learner can identify statements that imply cause and effect relationships in a paragraph.		(X)	(X)	X								
CI 10.0 DRAWING CONCLUSIONS - The learner can state logical conclusions for a reading selection.				(X)	(X)	(X)	(X)	(X)	X	X	X	X
10.1 The learner can identify logical conclusions about characters or events illustrated in a picture.	X											
10.2 The learner can identify logical conclusions about characters or events described in a story presented orally.	(X)	X										
10.3 The learner can identify logical conclusions about characters or events ina reading selection.		(X)	(X)	X	X	X						
CI 11.0 PREDICTING OUTCOMES - The learner can predict a logical outcome of a reading selection.				(X)	(X)	(X)	(X)	(X)	X	X	X	X
11.1 The learner can predict a logical outcome from a set of incomplete pictures.	X	X										
11.2 The learner can predict a logical outcome of a story presented orally.	(X)	X										
11.3 The learner can predict a logical outcome of a paragraph.		(X)	(X)	X	X							

DETAILS

GENERAL DESCRIPTION

The learner will be presented with an oral or written selection and asked to answer a question requiring the identification of a detail contained in that selection.

OVERVIEW OF OBJECTIVES TESTED

GRADE 1	GRADE 2	GRADE 3	GRADE 4	GRADE 5	GRADE 6	GRADE 7	GRADE 8
The learner can identify details in a story presented orally. (2.3)							
The learner can identify details in a reading selection. (2.4)							

25

26

DESCRIPTION OF TEST QUESTIONS

1. A test question will consist of either (a) an orally presented selection and question, or (b) a written selection and question.

2. The difficulty level of a selection will be determined by three factors. These are (a) the number of words in the selection, (b) the vocabulary level of the words, and (c) either the length of the sentences in the selection (Grades 1-3) or the readability rating of the selection as judged by the Fry readability formula (Grades 4-8). These factors will vary according to the grade level being tested as follows:

	GRADE 1	GRADE 2	GRADE 3	GRADE 4	GRADE 5	GRADE 6	GRADE 7	GRADE 8
Maximum length	50 words (oral)							
	35 words (written)	50 words	75 words	100 words	125 words	150 words	175 words	175 words
Maximum vocabulary level	Grade 3 word list (oral)							
	Grade 1 word list (written)	Grade 2 word list	Grade 3 word list	Grade 4 word list	Grade 5 word list	Grade 6 word list	Grade 7 word list	Grade 8 word list
Maximum sentence length	12 words (oral)							
	10 words (written)	12 words	14 words					
Fry readability rating				Grades 3-4	Grades 4-5	Grades 5-6	Grades 6-7	Grades 7-8